UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Simulated Action in an Embodied Construction Grammar

Permalink
https://escholarship.org/uc/item/6z0155f3

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 26(26)

ISSN
1069-7977

Authors
Bergen, Benjamin
Chang, Nancy
Narayan, Shweta

Publication Date
2004

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Simulated Action in an Embodied Construction Grammar

Benjamin Bergen (bergen@hawaii.edu)
Dept of Linguistics, 1890 East-West Hall, 569 Moore Hall
Honolulu, HI 96822

Nancy Chang (nchang@icsi.berkeley.edu)
Shweta Narayan (shweta@icsi.berkeley.edu)
International Computer Science Institute, 1947 Center St., Suite 600
Berkeley, CA 94704-1198, USA

Abstract

Various lines of research on language have converged on the premise that linguistic knowledge has as its basic unit pairings of form and meaning. The precise nature of the meanings involved, however, remains subject to the longstanding debate between proponents of arbitrary, abstract representations and those who argue for more detailed perceptuo-motor representations. We propose a model, Embodied Construction Grammar (ECG), which integrates these two positions by casting meanings as schematic representations embodied in human perceptual and motor systems. On this view, understanding everyday language entails running mental simulations of its perceptual and motor content. Linguistic meanings are parameterizations of aspects of such simulations; they thus serve as an interface between the relatively discrete properties of language and the detailed and encyclopedic knowledge needed for simulation. This paper assembles evidence from neural imaging and psycholinguistic experiments supporting this general approach to language understanding. It also introduces ECG as a model that fulfills the requisite constraints, and presents two kinds of support for the model. First, we describe two verbal matching studies that test predictions the model makes about the degree of motor detail available in lexical representations. Second, we demonstrate the viability and utility of ECG as a grammar formalism through its capacity to support computational models of language understanding and acquisition.

Introduction

Many theories of language take the basic unit of linguistic knowledge to be pairings of form and meaning, known as *symbols* or *constructions* (de Saussure 1916; Pollard & Sag 1994; Goldberg 1995; Langacker 1987). This view stems from the simple observation that language serves to convey meaning, using form. A speaker must thus know what linguistic forms are appropriate to encode the meanings s/he wishes to convey, and vice versa for an understander.

The nature of the meaning representations of linguistic units, however, remains very much at issue. Suggestions in the literature range from relatively abstract representations, including both feature structures (Pollard & Sag 1994) and logical representations (May 1985), to more concrete perceptual- or motor-based representations (Langacker 1987; Barsalou 1999; Glenberg & Robertson 2000).

Each of these approaches faces difficulties. Abstract symbol systems, whether feature-based or logical, invite the question of how (or even whether) they are ultimately linked to human perceptual, motor, affective, and other sorts of experience. There is strong evidence, seen below, that such embodied knowledge is automatically and unconsciously brought to bear during language understanding. Moreover, language users naturally make a broad range of associative and causal inferences based on language, a process not easily represented in an abstract symbol system.

Conversely, a theory of linguistic meaning cannot be based on perceptuo-motor information alone. Linguistic units can be combined in ways that are not strictly predictable from their semantic properties. Our ability to judge the grammaticality of sentences like Chomsky's (1957) classic *Colorless green ideas sleep furiously* example provides strong evidence of linguistic structure distinct from motor, perceptual, or other world knowledge. Additionally, our ability to understand sentences like *My pet chicken kissed me on the cheek* (even though chickens don't have lips, presumably a prerequisite for kissing) shows that grounded motor knowledge does not suffice to account for our ability to extract meaning from language.

One concrete solution to the drawbacks of purely abstract and purely perceptuo-motor approaches is to characterize mental representations as schematizations over modal knowledge (Fillmore 1982; Langacker 1987; Lakoff 1987; Barsalou 1999; Talmy 2000). This compromise view retains the best of both worlds: while language *use* involves the activation of perceptual and motor mechanisms, linguistic *units* themselves need only refer to schematic representations of these mechanisms. Proposals along these lines have inspired work investigating how the perceptual and motor structures underlying word meaning might be represented and schematized in computational models of human language processing (Regier 1996; Bailey 1997; Narayanan 1997). But the nature of the lexical and grammatical units that link these structures with linguistic forms has not yet been articulated precisely enough to support formal or computational implementation.

This paper synthesizes diverse evidence for an integrated view of language use and presents Embodied Construction Grammar (ECG), a formally specified instantiation of the approach. We begin by surveying evidence of the importance of perceptual and motor simulation in higher-level cognition, especially in language use. We then briefly outline the ECG formalism and show how it supports a model of human language use in which linguistic meanings serve to parameterize motor and perceptual structures. The remainder of the paper presents two kinds of support for the model. First, we describe a pair of verb matching studies that test predictions the model make about the degree of simulative detail in lexical representations. Second, we demonstrate the viability and utility of ECG as a grammar formalism precise enough to support computational models of language understanding and acquisition.

Mental Simulation in Language Use

Evidence for simulation

Perceptual and motor systems play an important role in higher cognitive functions, like memory and categorization (Barsalou 1999; Glenberg & Robertson 2000; Wheeler, Petersen & Buckner 2000; Nyberg et al. 2001), as well as motor (Lotze et al. 1999) and perceptual (Kosslyn, Ganis & Thompson 2001) imagery. It would thus be surprising if there were no role for perceptual and motor systems in language use as well.

Some theorists have proposed that perceptual and motor systems perform a central function in language production and comprehension (Glenberg & Robertson 2000; Barsalou 1999). In particular, they have suggested that understanding a piece of language entails internally simulating, or mentally imagining, the described scenario, by activating a subset of the neural structures that would be involved in perceiving the percepts or performing the motor actions described.

Several recent studies support this notion of simulation in language understanding, based on the activation of motor and pre-motor cortex areas associated with specific body parts in response to motor language referring to those body parts. Using behavioral and neurophysiological methods, Pulvermüller, Haerle & Hummel (2001) and Hauk Johnsrude & Pulvermüller (2004) found that verbs associated with different effectors were processed at different rates and in different regions of motor cortex. For example (Pulvermüller et al. 2001), when subjects perform a lexical decision task with verbs referring to actions involving the mouth (e.g. *chew*), leg (e.g. *kick*), or hand (e.g. *grab*), areas of motor cortex responsible for mouth/leg/hand motion displayed more activation, respectively. Tettamanti et al. (ms.) have also shown through an imaging study that passive listening to sentences describing mouth/leg/hand motions activates different parts of pre-motor cortex (as well as other areas, specifically BA 6, 40, and 44).

Behavioral methodologies also offer convergent evidence for the automatic and unconscious use of perceptual and motor systems during language use. Recent work on spatial language (Richardson et al. 2003; Bergen To Appear) has found that sentences with visual semantic components can result in selective interference with visual processing. While processing sentences that encode upwards motion, like *The ant climbed*, subjects take longer to perform a visual categorization task in the upper part of their visual field; the same is true of downwards-motion sentences like *The ant fell* and the lower half of the visual field. These results imply that language, like memory, evokes visual imagery that interferes with visual perception.

A second experimental method (Glenberg & Kashak 2002), tests the extent to which motor representations are activated for language understanding. The findings from this approach have shown that when subjects are asked to perform a physical action in response to a sentence, such as moving their hand away from or toward their body, it takes them longer to perform the action if it is incompatible with the motor actions described in the sentence. This suggests that while processing language, we perform motor imagery, using neural structures dedicated to motor control.

A third method, used by Stanfield & Zwaan (2001) and Zwaan et al. (2002), investigates the nature of visual object representations during language understanding. These studies have shown that implied orientation of objects in sentences (like *The man hammered the nail into the floor* versus *The man hammered the nail into the wall)* affected how long it took subjects to decide whether an image of an object (in this case, a nail) had been mentioned in the sentence, or even to name that object. It took subjects longer to respond to an image that was incompatible with the implied orientation or shape of a mentioned object. These results imply that not just trajectory and manner of motion, but also shape and orientation of objects, are represented in mental simulations during language understanding.

Linguistic knowledge as a simulation interface

Language understanding seems to entail the activation of perceptual and motor systems, which work in a dynamic, continuous, context-dependent, and open-ended fashion. Linguistic form, by contrast, is predominantly discrete — a word either precedes another word or does not; a morpheme is either pronounced or not, and so on. How do linguistic representations pair relatively discrete linguistic forms with continuous, dynamic, modal perceptuo-motor simulations? The notion of *parameterization* offers an answer.

Grammatical knowledge governing the productive combination of linguistic units appears to draw primarily on schematic properties of entities and events (Langacker 1987; Goldberg 1995), such as whether an entity can exert force or move, or whether an action involves the exertion of force or causes motion. Thus, for the purposes of language understanding, which involves determining what linguistic units an utterance uses and how they are combined, it may be sufficient for words and morphemes to generalize over perceptual or motor detail and encode only the important, distinctive aspects of actions and percepts required to perform a simulation. These parameterized representations are not abstract, amodal symbols, since they are directly grounded in action and perception, but they are distinct from the simulative content they parameterize.

This simulation-based view of language understanding has immediate consequences for theories of language. Only meaning representations that can be usefully fed to the simulation are viable; at the same time, constructions are freed from having to capture the encyclopedic knowledge handled by simulation. This division of labor between the meaning representations of linguistic constructions and the detailed structures that support simulation provides the means for finite, discrete linguistic structures to evoke the open-ended, continuous realm of possible meanings that language users may communicate. ECG is a theory of language that conforms to these constraints.

Embodied Construction Grammar

Embodied Construction Grammar (Bergen & Chang To Appear; Chang et al. 2002) aims to be a theory of language suitable for integration in a grounded, computationally implemented, simulation-based theory of human language use. It resembles other Construction Grammars (Kay & Fillmore 1999; Goldberg 1995; Croft 2001) in counting form-meaning pairings as the basic linguistic unit, and in aiming for full coverage of linguistic behavior. But ECG also serves as precisely the interface between language and simulation described above. It thus differs from other grammatical theories in emphasizing the embodiment of the grammatical system: constructions pair schematic form representations with schematic meaning representations, which are further constrained to be abstractions over perceptual and motor representations that can be simulated, or over characteristics of simulations in general.

A detailed description of the formalism is given in Bergen & Chang (To Appear); ECG has also been applied to a wide range of linguistic phenomena, including argument structure, reference, predication, and morphology, in a variety of languages. We concentrate here on showing how the representational tools of ECG satisfy and exploit the constraints of a simulation-based approach to language understanding. We first describe the high-level interactions posited in the model between linguistic constructions and the dynamic processes of language understanding they support, and then illustrate these with a simple example.

The Language Understanding Process

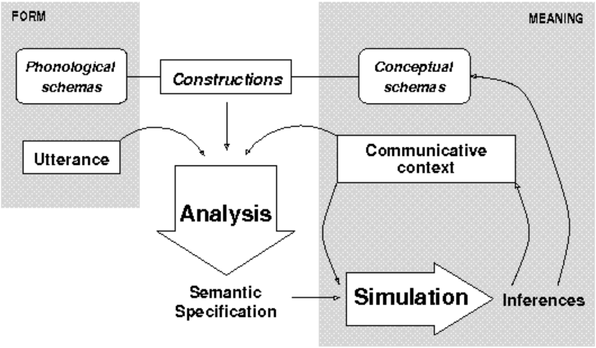

Figure 1. Language understanding in ECG.

The main source of linguistic knowledge in ECG is a large repository of constructions that express generalizations linking the domains of form (typically, phonological schemas and relations) and meaning (conceptual schemas and relations). Some constructions directly specify which perceptual and motor mechanisms to deploy, while others (especially larger phrasal and clausal constructions) specify how to combine the parameterized representations corresponding to different kinds of imagery. Still other constructions may affect the mode of simulation itself; the passive construction, for example, modulates what perspective is to be taken in the simulation of a given action.

There are also two main dynamic processes (large arrows in Figure 1) that interact with constructional knowledge during language comprehension. The first is the *analysis* process, which takes an input utterance in context and determines the set or sets of constructions most likely to be responsible for it. This process is thus roughly analogous to parsing, though it additionally incorporates contextual information, following Tanenhaus et al. (1995) and Spivey et al. (2001). The product of the analysis process is a structure called the *semantic specification* (or *semspec*), which specifies the conceptual schemas evoked by the constructions and how they are related. The second process is *simulation*, which takes the semspec as input and exploits representations underlying action and perception to simulate the specified events, actions, objects, relations, and states. The resultant inferences shape subsequent processing and provide the basis for the language user's response.

Embodied Construction Grammar in action

This section shows how the process just described would produce the appropriate simulation and resulting inferences for the sentence *Mary bit John*. The understander first tries to recognize the sequence of sounds in terms of form schemas. In speech or in sentences with novel or ambiguous word forms, this may require sophisticated categorization. Here, the forms are straightforwardly recognized as three form schemas ('Mary', 'bit', and 'John') with the appropriate temporal ordering relations among them, shown as vertical arrows on the left-hand side of Figure 2.

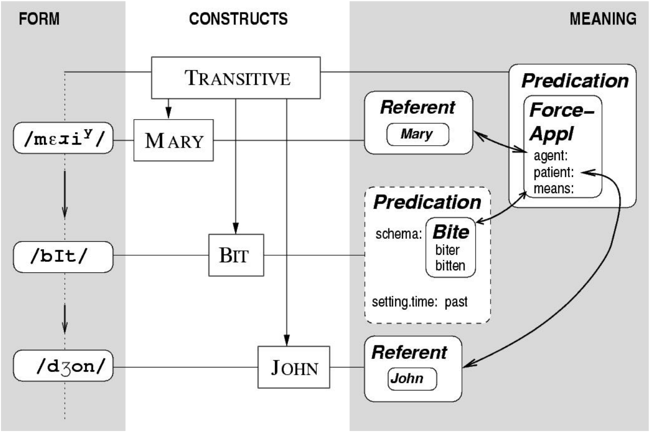

Figure 2: The (simplified) analysis of Mary bit John.

Next, the analysis process hypothesizes which constructions (instantiated as *constructs*) could account for the utterance; these are constructions whose form elements are present in the utterance. The four constructions relevant for this utterance are shown in the middle column of Figure 2. The JOHN and MARY constructions each bind some phonological form (on the left) with a particular schema for its referent (on the right). The BIT construction binds its phonological form with a predication that features a schema called Bite, which is the schematization of perceptual and motor knowledge about biting. Additionally, the predication is specified as being about the past, which will become relevant when inferences are propagated, after simulation.

The clausal TRANSITIVE construction binds together the forms and the meanings of the three lexical constructions, which serve as its constituents. On the form side, it specifies an ordering relation (Agent precedes Verb precedes Patient), while on the meaning side, the clause is linked to a predication that encodes the application of force by some means, where that means is bound to the Bite schema, the agent is Mary, and the patient is John. The clausal construction thus contributes its own structure and schematic meanings, and effects bindings among these and those of its constituents. In our example, the analysis process succeeds in finding a set of constructions that match the utterance and whose different constraints fit together, or unify, in all three domains: form, meaning and construction.

The completed analysis process produces a semspec (consisting of the meaning schemas and bindings, shown in the right-hand column in Figure 2), which is used as input for the next step, the mental simulation of the described scene. The semspec indicates which perceptual and motor structures should be activated and how they are related. It might also specify other parameters of the simulation, such as the perspective from which to simulate. Our example is in the active voice, not passive (e.g. *John was bitten by Mary*), so would by default be simulated from Mary's perspective, resulting in the activation of a motor schema for biting (though features of the surrounding context could result in an "experiencer" simulation perspective instead).

Although our example omits many details of analysis and simulation (including how the model supports, e.g., reference resolution, construal, and sense disambiguation (for a discussion of these, see Bergen and Chang (To Appear)), it nonetheless demonstrates how ECG captures the idea of parameterization or schematization. A verb form like 'bit' centrally includes a Bite schema in its meaning; this schema is a parameterization over perceptual and motor knowledge about biting. To figure out who is biting whom — that is, to understand how the meaning of 'bit' relates to the meanings of the other words in the utterance — only very general knowledge about biting (that it is an action in which force is exerted by one participant on another) is required. The simulation process makes use of the perceptual and motor knowledge underlying this schematic representation, and provides detailed perceptual and motor content that can support inference and, on the current account, constitutes understanding.

The remainder of the paper offers support for the ECG model from two different sources: a pair of behavioral experiments testing a prediction of the model, and the implementation of the formalism within computational models of language acquisition and understanding.

Experiment: Embodied Verbal Representation

The ECG approach to language claims that verbal semantics involves the activation of detailed motor knowledge about performing or perceiving the relevant action. One reflection of this prediction might be in the representation of the effector used to perform the action described by a verb. To test this hypothesis, we performed two related behavioral experiments. In the first, subjects were shown a picture and then a verb, and asked to decide whether the word correctly described the picture. In the second, subjects were asked to decide whether two verbs had nearly the same meaning. We predicted that subjects would take longer to reject as matches an image-verb or verb-verb pair that depicted different actions using the same effector, compared to the case when the two non-matches used different effectors.

Method

Study 1. 39 native English speakers participated for course credit. They were told that they would see a picture of a person performing an action on a screen, followed by a verb, and were instructed to decide as quickly as possible whether the verb was a good description of the picture.

During each trial, subjects were presented a stick figure of a person carrying out an action for 1 sec, a visual mask for 450 msec, and a blank screen for 50 msec. Then the verb was displayed, and stayed on the screen until the subject pressed "yes" or "no". All verbs were presented in the center of the screen. All actions were predominantly performed using one of three effectors: foot, hand or mouth. More detailed discussion of the methodology can be found in Bergen, Feldman & Narayan (2003).

Study 2. 53 native English speakers participated on a volunteer basis or for course credit. They were told that they would see a word appear on a screen and were instructed to decide as quickly as possible whether a second word that appeared meant more or less the same as the first word.

During each trial, subjects were presented with a fixation cross in the center of the screen for 2 sec, followed by an English action verb for 1 sec, a visual mask for 450 msec, and a blank screen for 50 msec. Then the second verb was displayed, and stayed on the screen until the subject pressed "yes" or "no". All verbs were capitalized and presented in the center of the screen. Verb pairs in critical trials pertained to motor actions of the following categories:

Matching: Near-synonyms, e.g.
SCREAM and SHRIEK; DANCE and WALTZ

Non-matching, same effector: Verbs with clearly different meanings, using the same effector, e.g.
SCREAM and LICK; DANCE and LIMP

Non-matching, different effector: Verbs with clearly different meanings, using different effectors, e.g.
SCREAM and STEP; DANCE and YELL

More detailed discussion of the methodology can be found in Narayan, Bergen & Weinberg (2004).

Results

Study 1. Counting only replies that were correct and within 2s.d. of the mean for a given subject, mean reaction times were 751ms for different-effector mismatches, 799ms for same-effector mismatches, and 741ms for matches. Using a standard ANOVA, the difference between the mismatching conditions was found to be significant ($p<.0001$).

Study 2. Counting only replies that were correct and within 3s.d. of the mean for a given subject, mean reaction times were 930ms for different-effector mismatches, 1030ms for same-effector mismatches, and 1070ms for near-synonyms.

Following Clark (1973), we performed two ANOVAs, with subjects and items as nested random factors, and from these determined that the RT difference between the mismatch conditions is significant (minF'(1,126)=9.0808, p<0.005). Post hoc tests showed that the non-matching different-effector condition is significantly different from the matching condition (minF'(1,126)=9.781), and the non-matching same-effector condition is not significantly different (minF'(1,126)=2.0002).

Discussion. Subjects took significantly longer to reject either a picture-verb pair as matches or a verb pair as near-synonyms when the two actions shared an effector than when they did not. Since this effect occurred when part of the task was non-linguistic (Study 1), this is unlikely to be a mere lexical effect. Moreover, the presence of the effect with purely linguistic stimuli (Study 2) means it is not due to strictly visual properties of the stimuli, either. Instead, these results suggest that understanding motion verbs goes beyond accessing abstract structures; modal information about bodily action, such as the effector used, is involved.

Importantly, the results imply that verb meaning does involve evoking modal motor representations: words encoding particular motor actions (kick, chew) contribute to the perceptuo-motor content of mental simulations.

ECG computational implementation

ECG is compatible in its broad outlines with a large body of linguistic and psycholinguistic research. But it is subject to the important additional constraint of being computationally precise. As we have described it, understanding even the simplest utterance involves multiple dynamic processes interacting with a variety of linguistic and embodied representations. Many of these are inspired by ideas from cognitive linguistics that have not been previously formalized, let alone used in any implemented system. It is thus crucial that we validate the framework by offering concrete implementations. In this section we briefly describe how the formalism serves as the lynchpin for computational models of linguistic use and acquisition.

Formally, the ECG construction and schema formalisms have much in common with other unification-based grammars (e.g., Pollard & Sag 1994), including notations for expressing features, inheritance, typing, and unification/coindexation; it also has additional mechanisms that increase its expressivity and flexibility.

As described earlier, the ECG formalism is designed to play a role in language understanding as the key interface between constructional analysis and the embodied simulation. Bryant (2003) describes an implemented construction analyzer that takes as input a grammar of ECG constructions and a sentence, and produces a semspec that provides the parameters for a simulation. The analyzer extends methods from syntactic parsing (particularly partial parsing and unification-based chart parsing) to accommodate and exploit the dual form-meaning nature of constructions. Specifically, it consists of a set of *construction recognizers*; each recognizer seeks the particular input form (or constraints) of its corresponding construction, and upon finding it checks the relevant semantic constraints. If multiple analyses are possible, the analyzer uses a *semantic density* metric to choose the analysis whose semspec is the most semantically coherent and complete. Thus, in constrast with typical language understanding systems in which syntactic parsing precedes semantic interpretation, the construction-based analyzer incorporates semantic constraints in parallel, reflecting the central role played by meaning in the ECG formalism.

The semspec produced by the analyzer provides parameters for simulation using active, modal structures. A broad range of embodied meanings have been modeled using *executing schemas* (x-schemas), a dynamic representation motivated in part by motor and perceptual systems (Narayanan 1997; Bailey 1997). X-schemas can model sequential, concurrent, and asynchronous events. The Bite schema, for example, parameterizes a Bite x-schema that captures the continuous mouth actions culminating in a particular forceful application of the teeth of the Biter to the Bitten. A simulation engine based on x-schemas has been implemented (Narayanan 1997) and applied to model the semantics of several domains, including verbal (Bailey 1997) and aspectual semantics (Chang, Gildea & Narayanan 1998), metaphorical inference (Narayanan 1999), and frame-based perspectival inference (Chang et al. 2002).

Although we have focused so far on language understanding, the ECG formalism is also designed to support a computational model of the acquisition of early phrasal and clausal constructions (Chang & Maia 2001; Chang 2004). This model takes ECG as the target representation to be learned from a sequence of utterances in context. Learning is usage-based in that utterances are first analyzed using the process described above; the resulting (partial) semspec is used along with context to prompt the formation of new constructions. The model has been applied to learn simple English motion constructions from a corpus of child-directed utterances, paired with situation representations. The resulting learning trends reflect cross-linguistic acquisition patterns, in particular the learning of lexically specific verb island constructions before more abstract grammatical patterns (Tomasello 1992). They also demonstrate how the ECG formalism serves as an interface between language comprehension and acquisition.

The implementations described here do not provide direct evidence of the cognitive reality of ECG. But they do demonstrate its utility and flexibility, and, by offering an integrated and precisely specified instantiation of simulation-based language understanding and use, serve as an existence proof for the overall approach.

Conclusions

If the embodied view presented above is correct, then the human capacity for language understanding relies on activating internal motor and perceptual simulations on the basis of linguistic input. These simulations can serve any of the purposes that linguistic information is conventionally put to — their content can be stored, thereby updating the internal knowledge base; their inferences can be propagated such that the understander can draw conclusions needed in discourse; or the actions they include can be performed in cases where the language involves instructions or requests.

The computationally viable and empirically supported model described above views linguistic units as pairings between schematic representations of form and schematic representations of meaning. Those representations are not abstract and arbitrary; rather, they are tightly bound to the perceptual and motor substrates over which they generalize. This approach permits insights into how language is integrated with perceptual and motor knowledge in the cognitive system, and sheds light on what meaning means.

Acknowledgements

Our thanks to Jerome Feldman, George Lakoff, Srini Narayanan, Zachary Weinberg, and other members of the NTL research group, as well as two anonymous reviewers for their helpful comments.

References

Bailey, D. 1997. When push comes to shove: A computational model of the role of motor control in the acquisition of action verbs. PhD thesis, UC Berkeley.

Barsalou, L. W. 1999. Perceptual symbol systems. Behavioral and Brain Sciences, 22, 577-609.

Bergen, B. & Chang, N. To Appear. Embodied Construction Grammar in Simulation-Based Language Understanding. In J-O. Östman and M. Fried (eds.) Construction Grammar(s): Cognitive and Cross-language dimensions. John Benjamins. http://www.icsi.berkeley.edu/NTL/papers/ecg-tr-02-004.pdf

Bergen, B. To appear. Experimental methods for simulation semantics. In M. Gonzalez-Marquez, I. Mittelberg, S. Coulson, and M. Spivey (eds.) Methods in Cognitive Linguistics: Ithaca.

Bergen, B., Narayan S., & Feldman, J. 2003. Embodied verbal semantics: evidence from an image-verb matching task. Proceedings Cognitive Science 25. Mahwah, NJ: Erlbaum.

Bryant, J. 2003. Constructional Analysis. UC Berkeley M.S. Report.

Chang, N, Gildea, D. & Narayanan, S. 1998. A Dynamic Model of Aspectual Composition. Proceedings Cognitive Science 20.

Chang, N. & T. V. Maia. 2001. Learning grammatical constructions. Proceedings Cog. Sci. 23. Mahwah: Erlbaum.

Chang, N., Feldman, J., Porzel, R., & Sanders, K. 2002. Scaling Cognitive Linguistics: Formalisms for Language Understanding. 2002. Proceedings of SCANALU.

Chang, N. 2004. Constructing Grammar: A computational model of the acquisition of early constructions. Ph.D. thesis, UC Berkeley.

Chomsky, N. 1957. Syntactic Structures. Mouton.

Clark, H. 1973. The language-as-fixed-effect fallacy. Journal of Verbal Learning and Verbal Behavior, 12:335–359.

Croft, W. 2001. Radical Construction Grammar. Oxford: Oxford University Press.

Fillmore, C. J. 1982. Frame semantics. In Linguistic Society of Korea (eds.), Linguistics in the Morning Calm.

Glenberg, A. M. & Kaschak, M. P. 2002. Grounding language in action. Psychonomic Bulletin & Review.

Glenberg, A. M. & Robertson, D. A. 2000. Symbol Grounding and Meaning: A Comparison of High-Dimensional and Embodied Theories of Meaning. JML, 43, 379-401.

Goldberg, A. 1995. Constructions: A Construction Grammar Approach to Argument Structure. Chicago: U Chicago Press.

Hauk, O., Johnsrude, I. & Pulvermüller, F. 2004. . Somatotopic representation of action words in human motor and premotor cortex. Neuron. 41(2): 301-7.

Kay, P. & Fillmore, C. J. 1999. Grammatical constructions and linguistic generalizations: The What's X doing Y? construction. Language 75/1: 1-33.

Kosslyn, S. M., Ganis, G. & Thompson, W. L. 2001. Neural foundations of imagery. Nature Reviews Neurosci., 2:635-642.

Lakoff, G. 1987. Women, fire, and dangerous things. Chicago: U Chicago Press.

Langacker, R. 1987. Foundations of Cognitive Grammar: Theoretical Perspectives 1. Stanford: Stanford University Press.

Lotze, M., Montoya, P., Erb, M., Hülsmann, E., Flor, H., Klose, U., Birbaumer, N., & Grodd., W. 1999. Activation of cortical and cerebellar motor areas during executed and imagined hand movements: An fMRI study. Jrn. Cog. Neurosci 11(5): 491-501

May, R. 1985. Logical Form. Cambridge: MIT Press

Narayan, S., Bergen, B., & Weinberg, Z. 2004. Embodied Verbal Semantics: Evidence from a lexical matching task. Proceedings of Berkeley Linguistics Society 30. Berkeley.

Narayanan, S. 1997 KARMA: Knowledge-based Action Representations for Metaphor and Aspect. Ph.D. thesis, UC Berkeley

Narayanan, S. 1999. Moving Right Along: A Computational Model of Metaphoric Reasoning about Events. Proceedings of the Nat. Conf. on Artificial Intelligence (AAAI '99): 121-128.

Nyberg, L., Petersson, K.-M., Nilsson, L.-G., Sandblom, J., Åberg, C., & Ingvar, M. 2001. Reactivation of motor brain areas during explicit memory for actions. NeuroImage, 14, 521-528.

Pollard, C. & Sag, I. 1994. Head-Driven Phrase-Structure Grammar. Chicago: University of Chicago Press.

Pulvermüller, F., Haerle, M., & Hummel, F. 2001. Walking or Talking?: Behavioral and Neurophysiological Correlates of Action Verb Processing. Brain and Language 78, 143–168.

Regier, T. 1996. The Human Semantic Potential: Spatial Language and Constrained Connectionism. Cambridge: MIT Press.

Richardson, D. C., Spivey, M. J., McRae, K., & Barsalou, L. W. 2003. Spatial representations activated during real-time comprehension of verbs. Cognitive Science.

de Saussure, F. 1916. Course de linguistique générale. Paris: Payot.

Spivey, M.J., Tanenhaus, M.K., Eberhard, K.M. & Sedivy, J.C. 2001. Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. Cognitive Psychology.

Stanfield, R. A. & Zwaan, R. A. 2001. The effect of implied orientation derived from verbal context on picture recognition. Psychological Science, 12, 153-156.

Talmy, L. 2000. Toward a Cognitive Semantics. Cambridge: MIT.

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. 1995. Integration of visual and linguistic information in spoken language comprehension. Science, 268, 1632-1634.

Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Perani, D., Cappa, S. F., Fazio, F., & Rizzolatti, G. Unpublished Ms. Sentences describing actions activate visuomotor execution and observation systems.

Tomasello, M. 1992. First verbs: A case study of early grammatical development. Cambridge: Cambridge University.

Wheeler, M. E., Petersen, S. E., & Buckner, R. L. (2000). Memory's echo: Vivid remembering reactivates sensory specific cortex. Proc. Natl. Acad. Sci. USA 97: 11125–11129.

Zwaan, R. A., Stanfield, R.A., & Yaxley, R. H. 2002. Do language comprehenders routinely represent the shapes of objects? Psychological Science, 13, 168-171.